

IoT 기반의 전력 모니터링 시스템에서 군집화와 군집 대표 부하를 활용한 무손실 데이터 압축 방법

이지훈, 윤승욱, 황의석*
광주과학기술원

{zoazoa61, ysw1207, *euisseokh}@gist.ac.kr

Lossless data compression based on clustering and cluster-representative load in IoT-enabled power monitoring system.

Lee Ji Hoon, Yoon Seung Wook, Hwang Eui Seok*
Gwangju Institute of Science and Technology (GIST)

요 약

본 논문은 사물 인터넷 (IoT, Internet of things) 환경에서 수집되는 전력 데이터의 압축을 목적으로 군집화 기술을 적용하여 군집 대표 부하를 활용한 무손실 압축 방법을 제안한다. 무손실 압축 방법은 원본 데이터의 손실 없이 복원할 수 있는 장점이 있어 중요한 정보의 보존을 위하여 활용되고 있다. 하지만, 개별 스마트 미터가 가지는 단일 데이터의 압축 방식은 압축률의 한계를 가지고 있어 이에 관한 효율적인 기술 개발이 중요하다. 제안된 압축 방식은 측정 장치로부터 수집된 전력 데이터의 프로파일 정보를 이용하기 위하여 군집화 기술을 적용하고 대표 및 비대표 부하로 나누어 유사 프로파일을 가진 데이터를 묶어 분류한다. 이후, 신호 간의 상관관계를 이용하여 선형 예측 부호화 기법을 적용함으로써 추정 신호를 생성한다. 이를 통해 원 신호와 추정 신호 간의 차이인 잔여 성분 신호를 생성하고 원 신호의 정보 엔트로피를 절감함으로써 단일 데이터의 압축 방식 대비 제안 방법을 적용하여 동일한 압축 알고리즘을 적용했을 경우 압축률이 향상됨을 확인하였다.

I. 서 론

최근, 스마트 가전과 사물 인터넷 (IoT, Internet of things)의 발달과 함께 관련 기술이 우리 일상생활에 적용됨에 따라 다양한 혜택이 제공되고 있다. 이에 따라 IoT 관련 기술의 기반으로 사용되는 전력 분석의 중요성 역시 증가하고 있다.[1] 예로서, 전력 데이터에 관한 많은 연구 중 비접촉식 개별 가전기기 부하 식별 (Non-intrusive load monitoring) 기술과 함께 측정 장비의 기술 수준이 높아지고 이에 따라 수집 데이터의 측정 간격을 줄이며 높은 샘플링을 가진 공개 데이터 세트의 크기가 점차 증가하였다. 이에 따라, 현재 상용화된 하드웨어의 사양은 제한적이므로 증가하는 데이터 세트의 효율적인 관리 문제는 중요해지고 있다. 따라서 데이터 세트의 크기를 효과적으로 압축하여 저장하고 압축 처리 속도를 줄이는 것이 필요하므로, 압축 알고리즘의 성능 향상에 관한 필요성이 제기되고 있다.[2]

압축 알고리즘은 크게 손실 압축과 무손실 압축 기술로 구분된다. 손실 압축 기술은 기존 정보에서 무의미한 정보에 관한 데이터를 제거한 후 압축하는 방법으로서, 높은 압축률을 얻는 것이 유리하여 이미지와 음원 데이터에 사용되고 있다. 반면, 무손실 압축 기술은 압축률 및 압축 속도와 비교하면 손실 압축보다 성능이 낮으나, 원본 데이터의 손실 없이 복원할 수 있는 장점이 있어 주로 텍스트 기반의 데이터와 같이 중요한 정보를 포함하는 데이터에 적용된다. 또한, 무손실 압축 알고리즘은 원본 데이터의 크기에 따라 계산 복잡도를 증가시키기 때문에 빅데이터의 처리 시, 압축 시간과 메모리 처리량

이 비례하여 증가하므로 무손실 압축 알고리즘의 성능 개선은 중요하다. 그러나 기존의 개별 스마트 미터가 가지는 단일 데이터 기반의 압축 방식은 압축률의 한계치를 가지고 있어 데이터의 그룹화를 통한 추가적인 압축률 향상에 관한 연구와 분석이 진행되고 있다.[3]

이를 위해, 본 논문에서는 전력 IoT 환경에서 수집되는 데이터 세트에 관하여 세부 그룹의 정보를 대표해서 개체 간 연관 관계를 설정하는 구조 [4]를 적용하고, 선형 예측 방법을 적용한 무손실 압축 알고리즘 [5]을 제안한다. 제안하는 방식은 스마트 미터 장치를 통해 수집되는 다양한 전력 데이터의 프로파일 정보를 이용하여 비지도 학습 기반 군집화 기술을 활용하고 대표 및 비대표 부하로 나누어 적용한다. 그리고 군집 내 대표 부하와 비대표 부하의 상관관계를 이용하여 선형 예측 모델 기반의 추정 신호를 생성한다. 이후, 원 신호와 추정 신호의 차이 값인 잔여 성분 신호를 저장하여 정보 엔트로피를 절감하고 기존 압축 알고리즘의 압축률 향상에 기여한다.

II. 군집 대표 부하를 활용한 무손실 데이터 압축 방법

본 논문에서 제안하는 압축 알고리즘은 그림 1 과 같이, 수집된 데이터에 군집화 기법을 적용한 후, 유사한 프로파일을 가진 데이터를 묶어 분류한다. 이후 군집 내 대표 부하와 비대표 부하를 선정하기 위해 유클리드 거리 (Euclidean distance)를 기준으로 군집의 중심 (Centroid)과 가장 가까운 데이터를 대표 부하로, 나머지 신호는 비대표 부하로 선정한다. 대표 부하를 기준으로

동일 군집 내 각 비대표 부하에 관하여 선형 예측 부호화 (LPC, Linear prediction coding) 기법을 적용하고 각각의 추정 신호를 생성한다. 이후, 원 신호와 추정 신호의 차이 값인 잔여 성분 신호를 생성하고 선형 계수 (Linear coefficient)와 함께 비대표 부하의 신호를 복원할 수 있는 정보로서 저장한다. 저장된 두 정보의 완벽한 복원을 위하여 무손실 압축 알고리즘을 적용함으로써 추가적인 압축률 향상을 기대한다.

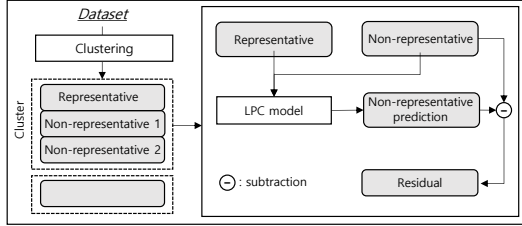


그림 1. 제안 방법의 구조도

III. 실험 환경

제안 방법의 검증을 위한 실험 환경은 다음과 같다. 실험 데이터는 공공 데이터 세트 중 하나인 Pecan street [6]로 정하여 14년도 9 개 주거형 데이터에 관한 1 분 단위 전력 데이터를 활용하여 실험을 진행하였다. 군집화 기법은 대표적으로 사용되는 K-평균 군집화 기법을 사용하였고 1 시간 단위 데이터의 중앙값을 기준으로 적용하였다. 최적 군집 수는 실루엣 기법 (Silhouette index)을 통해 4 개의 군집을 정하고, 군집의 중심 거리를 기준으로 가장 가까운 4526 번 집의 전력 정보를 대표 부하로, 나머지 4 개의 데이터를 비대표 부하로 선정하였다. 이후, 10 개의 선형 계수를 통해 LPC 기법을 적용하여 비대표 부하의 추정 신호를 생성하고 잔여 성분 신호를 구한 후 대표적인 무손실 압축 알고리즘인 LZMA, Bzip2, Gzip 등의 방법을 적용하여 압축을 진행하였다.[2]

VI. 실험 결과

제안 방법의 타당성 평가를 위해 먼저 (1)의 압축률을 적용하여 평가하고, 압축률과 정보 엔트로피 (2)의 상관 관계를 확인하며 이론적으로 실험 결과를 분석한다.

$$\text{Compressed ratio} = \left(\frac{\text{Compressed size}}{\text{Original size}} \right) * 100 \quad (1)$$

$$H(x) = -\sum_x P(x) \log P(x) \quad (2)$$

제안 방법을 적용한 각 비대표 부하의 압축률 평가 결과는 그림2와 같이 전체적으로 향상되었으며, LZMA 알고리즘을 적용했을 경우 가장 향상된 결과를 확인하였다. 또한, 표1과 같이 압축률과 정보 엔트로피와의 상관성을 확인할 수 있어, 잔여 성분 기반 제안 방법의 압축 성능 향상 효과를 정보 이론적 지표로 확인할 수 있다.

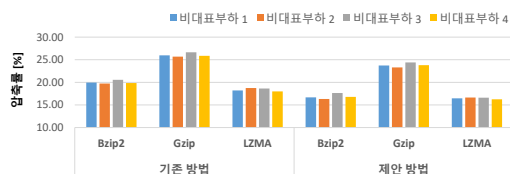


그림 2. 제안 방법에 관한 무손실 알고리즘 압축 평가

데이터	정보 엔트로피		압축률 [%]	
	원 신호	잔여 성분 신호	원 신호	잔여 성분 신호
비대표 부하 1	11.99	11.07	18.21	16.45
비대표 부하 2	12.94	11.88	18.74	16.62
비대표 부하 3	13.45	13.43	18.63	16.59
비대표 부하 4	11.07	10.56	17.98	16.25

표 1. 정보 엔트로피와 압축률의 시뮬레이션 결과

V. 결론

본 논문에서는 IoT 환경에서 개별 스마트 미터가 가지는 단일 데이터 기반의 압축 한계를 극복하기 위하여 군집화 기법을 적용하고 군집 대표 부하를 활용하는 무손실 압축 방법을 제안한다. 제안된 방법은 동일 군집 내 유사 프로파일을 묶어 분류한 후, 대표 및 비대표 부하의 상관관계를 이용하여 생성한 잔여 성분 신호를 압축함으로써 신호가 보유한 정보 엔트로피를 절감한다. 이를 통해 기존 단일 데이터에 관한 무손실 압축 알고리즘을 적용했을 경우 대비 제안하는 방법 적용 시 압축률이 향상됨을 실험을 통해 확인하였다.

ACKNOWLEDGMENT

이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 정보통신산업진흥원의 지원을 받아 수행된 에너지 AI 융합 연구개발 사업과 광주과학기술원의 GRI(GIST 연구원) 사업의 지원을 받아 수행된 연구임.

참 고 문 헌

- [1] Miraz M. "A review on internet of things (IoT), internet of everything (IoE) and internet of nano things (IoNT)," IEEE ITA, pp. 219-224, 2015.
- [2] Kriechbaumer T. "Waveform signal entropy and compression study of whole-building energy datasets," ACM ICFES, pp. 1-9, 2019.
- [3] Wang S. "A novel smart meter data compression method via stacked convolutional sparse auto-encoder," Electrical Power & Energy Systems, p. 105761, 2020.
- [4] Asimakopoulou G. "Leader-follower strategies for energy management of multi-microgrids," IEEE Trans. Smart Grid, pp. 1909-1916, 2013.
- [5] Wu J. "Lossless compression of hyperspectral imagery via clustered differential pulse code modulation with removal of local spectral outliers," IEEE Signal Processing Letters, pp. 2194-2198, 2015.
- [6] Pecan Street. "Dataport: the world's largest energy data resource," Pecan Street Inc, 2015.